

# **Estimating DGSE models with long memory dynamics**

G. Moretti and G. Nicoletti

Discussion Paper 2008-37

Département des Sciences Économiques  
de l'Université catholique de Louvain



---

**UCL**

---

# Estimating DSGE models with long memory dynamics

Gianluca Moretti and Giulio Nicoletti\*

## Abstract

Recent literature claims that key variables such as aggregate productivity and inflation display long memory dynamics. We study the implications of this high degree of persistence on the estimation of Dynamic Stochastic General Equilibrium (DSGE) models. We show that long memory data produce substantial bias in the deep parameter estimates when a standard Kalman Filter-MLE procedure is used. We propose a modification of the Kalman Filter procedure, we mainly augment the state space, which deals with this problem. By the means of the augmented state space we can consistently estimate the model parameters as well as produce more accurate out-of-sample forecasts compared to the standard Kalman filter.

## 1 Introduction<sup>1</sup>

There is a widespread consensus that key macroeconomic variables such as aggregate productivity and inflation are characterized by very persistent dynamics. Recent

---

\*Bank of Italy, Research Department

<sup>1</sup>We thank Raf Wouters for discussing with us a preliminary version of the paper. We are indebted to Stefano Neri, F. Mihoubi and participants at the ‘Third Dynare Conference’ for helpful comments at various stages of the paper. We are also grateful to Fabio Buseti, Michele Caivano, Sergio Santoro, Enrico Sette and Giordano Zevi for useful comments. The usual disclaimer applies.

empirical literature argues that the dynamics of most macroeconomic data can be better represented by long memory processes, specifically fractionally integrated, rather than standard autoregressive moving average (ARMA). Diebold and Rudebusch (1989) pioneered the introduction of long memory processes to describe U.S. GDP. The same result was recently confirmed by Mayoral (2005) using an updated version of their dataset. Recent studies by Gadea and Mayoral (2005) and Altissimo et al. (2005) show evidence of long memory in the inflation of the OECD countries and the euro area; while Abadir et. al. (2006) documents analogous behavior for the Nelson and Plosser database.

In this paper we analyze the implications of a high degree of persistence in the data on the estimation of Dynamic Stochastic General Equilibrium (DSGE) models. First we show that long memory can give rise to substantial bias in the estimates of the deep parameters of the model. Second we propose an approach to effectively tackle this problem.

Over the last few years, DSGE models have become the workhorse of modern macroeconomic modelling, moreover both academics and practitioners use them to produce macroeconomic forecasts. The reduced form of these models (i.e. what is obtained once they are solved for expectations) describe the data as a (dynamic) linear combination of fundamental shocks and endogenous states. The reduced form of the model can be cast into a state space form. It has been shown that under some general conditions this implies a finite VARMA representation for the data. The order of the MA component depends in general upon the number of endogenous states (such as capital) which are treated as non-observed variables.<sup>2</sup>

According to the theoretical literature, long memory arises from the aggregation of heterogeneous ARMA processes (see Granger (1980) and Chambers (1998)): this procedure generates an aggregate process which is much more persistent than its underlying components. Moreover, as Granger and Joyeux (1980) showed, the ag-

---

<sup>2</sup>The order of the VARMA depends upon the type of model at hand. In particular the order of the AR component depends upon the order of the AR in the exogenous states while the MA might depend upon the presence of endogenous states which are treated as non observed variables, see Ravenna (2007) for more details.

aggregate dynamics can not be adequately captured by finite order ARMA processes<sup>3</sup>. This implies that the finite order VARMA representation of DSGE models could be misspecified when faced to long memory data; in particular, as we show below, it can lead to biased parameter estimation as well as less accurate forecasts. A likely way to deal with the problem would be to introduce a large number of unobserved endogenous states in the model to increase the order of the VARMA representation. Although this approach might improve the fit of the model, it does not really tackle the issue of possible long memory in the data.

A direct attempt to introduce long memory dynamics in a DSGE model has been recently proposed by Abadir and Talmain (2002). Within a monopolistic competition framework, they show that heterogeneity in the firms' technology generates aggregate dynamics for the output that are consistent with the shape of the autocorrelation function for the U.S. GDP. Despite the novelty of the result, their approach is computationally very demanding and cannot be easily extended to models more complex than the one considered in their paper. This highlights a key problem: complex dynamics and better approximation of persistent data comes at the cost of an increased complexity and high computational burden. Our aim is to propose an approach that reconciles the dynamic properties of the observed data with the stylized representation of DSGE model.

The contribution of this paper is twofold. First, we show that a strong degree of persistence in the data can substantially bias the structural parameters estimates of a DSGE model. Following the thread of Mcgrattan (2007) and Ruge-Murcia (2007) we simulate artificial data samples from a Real Business Cycle (RBC) model where technology shocks are generated from an ARFIMA process consistently with the evidence of Diebold and Rudebusch (1980) and Sowell (1992). We evaluate the ability of the Maximum Likelihood (ML) methods to estimate the DSGE structural

---

<sup>3</sup>A similar result is discussed recently by Carvalho (2007) where he shows that there is no simple link between a new keynesian model with heterogeneity in the Calvo pricing frequency and one with a single representative frequency of Calvo adjustment. While Carvalho focuses on impulse responses and one specific calibrated example of aggregation, here we propose a general method to bring 'representative' models to long memory data, sidestepping the specific aggregation problem at hand.

parameters for different degrees of the persistence. We find a relevant amount of bias in the estimates: the stronger the persistence of the data the larger the bias.

Second, we propose an approach that copes with the strong persistence in the data within a stylized DSGE model. Specifically we develop an ‘augmented’ state space<sup>4</sup> which is consistent with the hypothesis of long memory in the exogenous state processes; we show through simulation that our method is able to consistently estimate the real parameters of the model over all the possible degrees of persistence in the data generating process.<sup>5</sup>

We then take our augmented Kalman filter procedure to the real data, and following Ireland (2004), we estimate a standard RBC model on U.S. data. We report significantly different and more plausible parameter estimates compared to the standard maximum likelihood approach. Moreover, in order to assess the ability of the augmented Kalman filter to better capture the dynamics of the data, we also evaluate its out of sample forecast accuracy against the standard one. The prediction of the augmented filter outperforms significantly (by the means of a Diebold-Mariano test) those of the standard filter: we report an average reduction in the forecast error of about 30%. This result is consistent with the findings of Granger and Joyeux (1980) that even if autoregressive models can fit long memory dynamics in finite samples, they produce quite inaccurate out of sample forecasts.

The plan of the work is as follows. In section 2 we describe the modified Kalman filter and the maximum likelihood estimation of its parameters when the degree of memory of the data is unknown. In section 3 we run some simulation to evaluate the bias in the estimation of a RBC model for different degrees of persistence in the underlying data. We then repeat the same exercise for the estimation approach proposed in section 2. In section 4 we estimate a RBC model using real data for the U.S. economy. Furthermore, we evaluate the forecast accuracy of the two approaches. Section 5 concludes.

---

<sup>4</sup>In the text we refer to our procedure as ‘augmented Kalman filter (procedure)’ or to ‘augmented state space’.

<sup>5</sup>When we refer to strong or high degree of persistence we exclude the case of unit root. By definition, unit root implies infinite memory and therefore it can not fall into the case considered here.

## 2 Methodology

In this section we develop the augmented state space and describe how to implement it to estimate the parameters of a DSGE model. In section 2.1 we present the equations of the augmented filter assuming that the unknown exogenous process is a long memory process and that the autocorrelation structure of the data is known. In section 2.2, we remove this latter hypothesis and describe our estimation strategy.

### 2.1 Augmented State Space

As we see in the next section, the solution of many DSGE models is usually represented in a state-space form, i.e.

$$\theta_{t+1} = \Phi\theta_t + \varepsilon_{t+1} \quad (1)$$

$$y_t = H\theta_t + v_t \quad (2)$$

$$E(\varepsilon_t \varepsilon'_{t+k}) = \begin{cases} Q & k = 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$E(v_t v'_{t+k}) = \begin{cases} R & k = 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $\theta_t$  is the vector of the state and exogenous variables which evolve as an autoregressive process (AR) with innovation vector  $\varepsilon_t$ ;  $y_t$  is the vector of observed variables and  $v_t$  is the vector of measurement errors. The evolution path of the state variables is usually unknown, but under the assumption that it has an AR representation as in eq.1, it can be estimated using the Kalman Filter. Specifically, if we define respectively with

$$\begin{aligned} \theta_{t|t-1} &= E(\theta_t | \mathfrak{S}_{t-1}) \\ P_{t|t-1} &= E[(\theta_t - \theta_{t|t-1})(\theta_t - \theta_{t|t-1})'] \end{aligned}$$

the optimal estimator of the state  $\theta_t$  based on all the information up to  $t - 1$  and its dispersion matrix, then the Kalman filter is represented by the following equations,

$$\theta_{t|t-1} = \Phi\theta_{t-1} \quad (5)$$

$$P_{t|t-1} = \Phi P_{t-1} \Phi' + Q \quad (6)$$

$$F_t = E \left[ (y_t - H\theta_{t|t-1}) (y_t - H\theta_{t|t-1})' \right] = H P_{t|t-1} H' + R \quad (7)$$

$$\theta_t = \theta_{t|t-1} + K_t (y_t - H\theta_{t|t-1}) \quad (8)$$

$$K_t = P_{t|t-1} H_t' (H P_{t|t-1} H' + R)^{-1} \quad (9)$$

$$P_t = (I - K_t H) P_{t|t-1} \quad (10)$$

A detailed exposition on the derivation of the Kalman filter goes beyond the purpose of this paper and it can be found in Hamilton (1994). However, it is worth recalling that eq.5 and eq.6 are respectively the state prediction<sup>6</sup> and its variance, given the information set at  $t-1$ . Equation 7 is the dispersion matrix of the prediction error of  $y_t$ , given the information available at time  $t - 1$ . The matrix  $K_t$  is the so-called Kalman gain and it is a weighting matrix that minimizes the variance of the state forecast error. In other words, it solves the following problem:

$$\min_K E \left[ (\theta_t - \theta_{t|t-1} - K_t (y_t - H\theta_{t|t-1})) (\theta_t - \theta_{t|t-1} - K_t (y_t - H\theta_{t|t-1}))' \right]$$

The Kalman gain is the relative weight given to the observed variable in forming the prediction of the state variable  $\theta_t$  at time  $t$ . This weight is negatively related to the variance of the measurement error  $R$ : the larger is  $R$ , the smaller is the Kalman Gain  $K_t$  and the less importance is given to the measurement error when making the forecast of the state at time  $t$ , namely  $\theta_t$ , given the information set at time  $t - 1$ . Finally,  $P_t$  is the variance covariance matrix of the state conditional on information at time  $t$ .

---

<sup>6</sup>To ease the notation we denote with  $\theta_t$  the prediction of the state  $\theta_t$  at time  $t$  given the information set at time  $t$ , i.e.  $\theta_{t|t}$ .

Equations 5 and 10 represent a system whose parameter can be estimated by maximum likelihood. In fact, if we assume that the innovations  $\{e_t\}$  and  $\{v_t\}$  are multivariate Gaussian, then the conditional distribution of  $y_t$ , given the state  $\theta_t$  and the information at time  $t - 1$ , is given by

$$f_{y_t|\theta_t, \mathfrak{S}_{t-1}}(y_t|\theta_t, \mathfrak{S}_{t-1}) = (2\pi)^{-\frac{1}{2}} \left| HP_{t|t-1}H' + R \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y_t - C\theta_{t|t-1}) \left( HP_{t|t-1}H' + R \right)^{-1} (y_t - C\theta_{t|t-1})' \right\} \quad (11)$$

which can be maximized with respect to the unknown parameters.

The Kalman filter representation above is based on a number of assumptions about the data generating process (DGP). It assumes that the unobserved state variables evolve as autoregressive processes; the variance-covariance structure of the innovations  $\varepsilon_t$  and  $v_t$  is known; the relations between the state and the observed variables are linear. The last two technical assumptions can be somehow relaxed,<sup>7</sup> but, since the true DGP is unknown, the first is rather arbitrary and can affect the estimation results. In other words, a correct parameter estimation would require a correct specification of the dynamics of the unobserved variables: if data are characterized by strong persistence, then choosing too few lags in the AR representation of the state could leave a substantial amount of autocorrelation in the residuals: this leads to uncorrect prediction of the state dynamics which in turn leads to bias in the estimated parameters. An intuition on the problem of state prediction can be provided in the following way. The Kalman Filter is based on two steps: a projection step (as in equation 5), where states are projected ahead using the transition equation, and an information updating step (see equation 8) when the state estimates are revised due to new data arrival: with long memory both stages are affected by the misspecification. The former one is directly influenced since the one-step-ahead prediction directly depends on the specified transition equations. On top on that the information updating step is a simple sum of previous state estimate plus the innovation weighted by the Kalman gain, i.e.:

---

<sup>7</sup>See Durbin and Koopmans (2001)



$$\theta_{t|t} = \theta_{t|t-1} + K\nu_t ; \nu_t \equiv y_t - H\theta_{t|t-1},$$

for this to be optimal the innovations  $\nu_t$  generated by the Kalman Filter at time  $t$ , are assumed to be orthogonal to the estimated state based on information at  $t-1$ ,  $\theta_{t|t-1}$ . This need not be true when there are too few lags in the AR representation of the exogenous states, leaving a substantial amount of autocorrelation in the residuals.

In the next few paragraphs, we describe our approach to estimating DSGE models under the assumption that their dynamics are driven by this kind of process. To make the procedure more comprehensible we focus on the case where the state vector is composed by only one exogenous process. Specifically, if we recall the state-space defined in eq.-5 - 10,

$$\begin{aligned} \theta_{t+1} &= \Phi\theta_t + \varepsilon_{t+1} \\ y_t &= H\theta_t + v_t \\ E(v_t v'_{t+k}) &= \begin{cases} R & k = 0 \\ 0 & otherwise \end{cases} \end{aligned}$$

we consider the case when the innovation  $\{\varepsilon_t\}$  is a fractional noise, i.e.

$$(1 - L)^d \varepsilon_t = e_t \tag{13}$$

This implies that the exogenous variable  $\theta_t$  belongs to the class of fractional autoregressive processes, i.e.  $(1 - \Phi L)(1 - L)^d \theta_{t+1} = \varepsilon_{t+1}$ . This kind of process has been extensively studied in time series analysis and a good review of their properties can be found in Robinson (1994). In the next paragraphs we recall a few properties which are related to the degree of persistence of this kind of process. First, if  $d = 0$ , then  $\varepsilon_t$  is a standard white noise process, while if  $d = 1$  then  $\varepsilon_t$  is a random walk. If  $0 < d < 0.5$  the process is stationary but is said to have long memory in the sense of hyperbolic rather than exponential decay of the autocorrelation function. Finally, if  $0.5 < d < 1$  the process is non stationary, but differently from unit root

processes, still mean reverting. A relevant feature of this process is that for  $d < 0.5$  the autocorrelation function  $\rho_\varepsilon(k)$  of  $\varepsilon_t$  decays at the rate

$$\rho_\varepsilon(k) \simeq Ak^{2d-1}$$

which is somewhat slower than the exponential decay of the autocorrelation of AR processes. Furthermore, since

$$(1 - L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)(-1)^j}{\Gamma(j+1)\Gamma(-d)} L^j \simeq \sum_{j=0}^{\infty} j^{-(j+1)} \frac{(-L)^j}{\Gamma(-d)}$$

where  $\Gamma(\cdot)$  is the Gamma function, then the innovation  $\varepsilon_t$  has an infinite order autoregressive representation  $\text{AR}(\infty)$  given by

$$(1 - L)^d \varepsilon_t \simeq \sum_{j=0}^{\infty} j^{-(j+1)} \frac{(-L)^j}{\Gamma(-d)} \varepsilon_t = \sum_{j=0}^{\infty} \Pi_j \varepsilon_{t-j} + e_t$$

This implies that if the data is generated by eq.13, choosing the right number of lags in the representation of the state 1 is crucial to get consistent estimates of the model parameters: too few lags could in fact bias the parameters estimates due to possible autocorrelation left over in the residuals. In the next paragraph we describe how to construct a filter (unfeasible filter) which is consistent with an observed component as defined in eq.13. For this purpose, we start by assuming that the correlation matrix of  $\varepsilon_t$  is known. We will relax this assumption in the next section where we describe how to estimate the model parameters from the data. The filter we propose accounts for dynamic miss-specification in the state process and follows the idea brought forward by Abadir and Talmain (2006) and further developed in Moretti (2007). The ratio behind this approach relates to a Generalized Least Square (GLS) type correction to clean any left over autocorrelation in the estimation procedure. We start by defining

the variance-covariance matrix of  $\varepsilon_t$  up to lag  $m$ , namely  $E(\varepsilon_t \varepsilon_{t-k}') = \Omega^m$ , as

$$\Omega^m = \sigma^2 \begin{bmatrix} 1 & \rho_\varepsilon(1) & \cdots & \rho_\varepsilon(m) \\ \rho_\varepsilon(1) & 1 & \cdots & \rho_\varepsilon(m-1) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_\varepsilon(m) & \rho_\varepsilon(m-1) & \cdots & 1 \end{bmatrix}$$

where  $\rho_\varepsilon(k)$  is the  $k$ -th autocorrelation and  $\sigma^2$  is variance of  $\varepsilon_t$ . We consider the Cholesky decomposition of  $\Omega^m$ , namely

$$\Omega^m = \sigma^2 \Gamma \Gamma'$$

where  $\Gamma$  is lower triangular with elements  $\{\gamma_{i,j}\}_{i,j=1}^{m+1}$ . The ratio behind a GLS procedure is based on the construction of a vector of transformed variables  $\mathbf{z}_t$  defined as

$$\mathbf{z}_t = \begin{bmatrix} z_{t-m} \\ \vdots \\ z_{t-1} \\ z_t \end{bmatrix} = \begin{bmatrix} l_{0,1} & 0 & \cdots & 0 \\ l_{1,1} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ l_{m,1} & l_{m,2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \theta_{t-m} \\ \vdots \\ \theta_{t-1} \\ \theta_t \end{bmatrix} \equiv L \theta_t \quad (14)$$

where  $L \equiv \Gamma^{-1}$ , in order to remove any autocorrelation in the regression residuals in eq.1. This implies that, by the definition in eq. 1, the transformed variable  $z_t = \theta_t + \sum_{j=0}^{m-1} l_{m,m-j} \theta_{t-j-1}$  is an AR(1) process, i.e.

$$z_{t+1} = \Phi z_t + e_{t+1} \quad (15)$$

The coefficient  $l_j$  can be thought of as optimal weights such that  $E(e_t | z_t) = 0$  is satisfied when regressing  $z_t$  on its lagged value. This is the implicit moment condition that we will impose in the next section when estimating the parameters of the model. It can be immediately seen that if the  $\varepsilon_t$  are uncorrelated, then the coefficient  $\{l_{m,m-j}\}_{j=0}^{m-1}$  are all equal to zero<sup>8</sup> and  $\theta_t$  is equal to  $z_t$  which leads us

---

<sup>8</sup>since  $\Gamma \Gamma' = I$

back to the standard case. We are now able to define what we call henceforth the Augmented State Space model.

**Definition 1** *Lets consider the variables defined in the state space model in eq.1 -2, with  $\varepsilon_t$  being a fractional noise as defined in eq.13 and  $v_t$  white noise with variance  $\sigma^2$ , then the augmented state-space model is defined as*

$$\begin{bmatrix} z_{t-m+1} \\ \vdots \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{(m)} & I_{(m)} \\ \mathbf{0}'_{(m)} & \Phi \end{bmatrix} \begin{bmatrix} z_{t-m} \\ \vdots \\ z_t \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{(m)} \\ e_{t+1} \end{bmatrix} \quad (16)$$

or

$$\begin{aligned} \mathbf{z}_{t+1} &= \Psi \mathbf{z}_t + \mathbf{e}_{t+1} \\ \theta_t &= z_t + \sum_{j=0}^{m-1} \gamma_{m,m-j} z_{t-j-1} = \begin{bmatrix} \mathbf{0}_{m-1} & 1 \end{bmatrix}' \Gamma z_t = D_m' \Gamma z_t \end{aligned} \quad (17)$$

$$y_t = H\theta_t + v_t \quad (18)$$

$$E(\mathbf{e}_t \mathbf{e}_t') = \left\{ \sigma^2 D_m D_m' = \tilde{Q} \right. \quad (19)$$

$$E(v_t v_t') = \{R \quad (20)$$

where  $\mathbf{0}_{(m)}$  is a zero vector of  $m$ -elements and  $\gamma_{m,m-j}$  correspond to the  $m, m-j$  elements of the matrix  $\Gamma$ .

The difference with the standard state-space models is given by eq. 17 which can be considered as a “bridge variable” that embodies all the information on the autocorrelation function of  $\varepsilon_t$ . Once again, if the  $\varepsilon_t$  are uncorrelated (i.e.  $d = 0$ ), then the Augmented state space model reduces to the standard state space model.<sup>9</sup> The Kalman filter equations for the state-space model defined in eq.16-20 are given

---

<sup>9</sup>The filter by construction requires that we drop the first  $m$  observation in order to avoid any effect of the initial condition at the beginning of the sample.

by

$$z_{t|t-1} = \Psi z_{t-1} \quad (21)$$

$$P_{t|t-1}^z = \Psi P_{t-1}^z \Psi' + \tilde{Q} \quad (22)$$

$$\theta_{t|t-1} = z_{t|t-1} + \sum_{j=0}^{m-1} \gamma_{m,m-j} z_{t-j-1|t-1} = D_m' \Gamma z_{t|t-1} \quad (23)$$

$$\theta_t - \theta_{t|t-1} = (z_t - z_{t|t-1}) + \sum_{j=0}^{m-1} \gamma_{m,m-j} (z_{t-j} - z_{t-j|t-1}) = D_m' \Gamma (z_t - z_{t|t-1}) \quad (24)$$

$$P_{t|t-1}^\theta = P_{t|t-1}^z + \sum_{j=0}^{m-1} \gamma_{m,m-j}^2 P_{t-j-1|t-1}^z = D_m \Gamma' P_{t|t-1}^z \Gamma D_m' \quad (25)$$

$$\begin{aligned} F_t &= E \left[ (y_t - H\theta_{t|t-1}) (y_t - H\theta_{t|t-1})' \right] = H P_{t|t-1}^\theta H' + R \\ &= H D_m \Gamma' P_{t|t-1}^z \Gamma D_m' H' + R \end{aligned} \quad (26)$$

$$z_t = z_{t|t-1} + K_t (y_t - D_m' \Gamma z_{t|t-1}) \quad (27)$$

$$K_t = P_{t|t-1}^z H_t' D_m' \Gamma (H D_m \Gamma' P_{t|t-1}^z \Gamma D_m' H' + R)^{-1} \quad (28)$$

$$P_t^z = (I - K_t H) P_{t|t-1}^z \quad (29)$$

The augmented filter differs from the standard Kalman filter in the following ways. First, the block of equations 23 - 25 links the transformed exogenous state to the observed variables through the weighting matrix  $\Gamma$ . If the  $\varepsilon_t$  are uncorrelated then  $\Gamma = I$  and  $\theta_{t|} = z_t$ ; we are back to the standard Kalman filter. The variance of the state prediction error  $P_{t|t-1}^\theta$  in eq.27 depends on matrix  $\Gamma$  and it is generally larger than the matrix  $P_{t|t-1}$  defined in the standard Kalman filter (6). This implies that also the Kalman gain  $K_t$  defined in 28 will be larger than its equivalent in the standard Kalman filter. This is because the augmented filter embodies all the information contained in the autocorrelation function of the state  $z_t$  and consequently gives more weight (compared to the standard filter) to the observed variable when estimating the state variable at time  $t$  given the information at time  $t - 1$ . On the other hand, the standard Kalman filter, by imposing a specified AR formulation to

the state dynamics, would discard all the residual persistence and regard it as a noise component of the observed variable. As we show below in the simulation exercise this results in a smaller Kalman gain and a more “prominent role” of the past prediction of the states  $z_{t|t-1}$  in forming the new estimate of the state  $z_t$ .

## 2.2 Estimation

In this section, we consider the case when the autocorrelation matrix  $\Omega^m$  is unknown and describe how to estimate it together with the model’s parameters. The rationale behind our approach can be thought of as a generalized method of moments estimation with a “GLS type” correction to clean for eventual autocorrelation in the estimation residuals.

In the previous paragraphs we assumed that the elements of  $\Omega^m$  were known. Generally, they are not known and apart from very few special cases it is not possible to estimate  $\Omega^m$  since we would have to estimate a very large number of parameters. A solution to this problem has been proposed by Abadir and Talmain (2007) who suggested to fit the ACF of  $\varepsilon_t$  using the functional form

$$\rho_\varepsilon(k) \simeq \frac{1}{(1 + a_1 k^{a_2})^{a_3}}$$

where  $a_1$ ,  $a_2$  and  $a_3$  are parameters to be estimated. This functional form<sup>10</sup> was derived in Abadir and Talmain (2002) and corresponds to the decay rate of the ACF of a long memory process which includes as a special case the fractional integrated processes<sup>11</sup>.

---

<sup>10</sup>This functional form has been used in a number of papers. Abadir et. al. showed that it can capture very closely (and better than ARMA processes) the dynamic properties of many economic variables; Abadir and Talmain (2006) used it to construct a GLS approach, similar to the one proposed here, and to solve the uncovered interest rate puzzle; Moretti (2007) used it to develop a test for long memory co-movements between two macroeconomic time series.

<sup>11</sup>A different way of proceeding would be to impose a specific long memory functional form, estimate the  $d$  parameter and use that information in the estimation. While this approach might fall back into a misspecification problem, we would not gain any further information from doing that since our objective is to clean residuals from the autocorrelation. The  $d$ , which we do not see as a structural parameter, could be anyway recovered from the formula above.

Therefore, if we assume that the innovations  $\mathbf{e}_{t+1}$  and  $v_t$  are normally distributed, then we can estimate the parameters of the augmented filter by maximizing the following likelihood function

$$\begin{aligned}
f_{y_t|\theta_t, \mathfrak{S}_{t-1}}(y_t|\theta_t, \mathfrak{S}_{t-1}) &= (2\pi)^{-\frac{1}{2}} \left| HP_{t|t-1}^\theta H' + R \right|^{-\frac{1}{2}} \\
&\exp \left\{ -\frac{1}{2} (y_t - H\theta_{t|t-1}) \left( HP_{t|t-1}^\theta H' + R \right)^{-1} (y_t - H\theta_{t|t-1})' \right\} \\
&= (2\pi)^{-\frac{1}{2}} \left| HD\Gamma P_{t|t-1}^z \Gamma' D' H' + R \right|^{-\frac{1}{2}} \\
&\exp \left\{ -\frac{1}{2} (y_t - HD'\Gamma' z_{t|t-1}) \left( HD\Gamma P_{t|t-1}^z \Gamma' D' H' + R \right)^{-1} (y_t - HD'\Gamma' z_{t|t-1})' \right\}
\end{aligned}$$

where

$$\begin{aligned}
\Gamma\Gamma' &= \sigma^2 \Omega^m; \\
\Omega^m &= \{\omega_{i,j} : \omega_{i,j} = \rho_\varepsilon(k), \ k = |i-j|, \ i, j = 1, \dots, m\} \\
\rho_\varepsilon(k) &= \frac{1}{(1 + a_1 k^{a_2})^{a_3}}
\end{aligned}$$

with respect to the parameters of the filter matrices and the parameters of the functional form.

This procedures is equivalent to minimizing the quantity

$$Q(y_t|\theta_t, \mathfrak{S}_{t-1}) = \left( y_t - D'\Gamma' z_{t|t-1} \right) \left( HD\Gamma P_{t|t-1}^z \Gamma' D' H' + R \right)^{-1} \left( y_t - D'\Gamma' z_{t|t-1} \right)'$$

which implicitly requires the moment condition  $E(e_t|z_t) = 0$  to be satisfied. Therefore, our approach can be seen as a generalized method of moments estimation with a “GLS type” correction where  $\Gamma'$  is an optimal weighting matrix such that  $e_t$  and  $z_t$  are uncorrelated. In the next section, we show through simulation the ability of our approach to estimate consistently the structural parameters of a simple DSGE model under the hypothesis that the DGP of the state is a fractional AR process.

### 3 The artificial economy

In this section we assess through simulation the effects of long memory on the estimation of a DSGE model. For this purpose, we simulate an aggregate RBC economy with a single exogenous shock where we assume that the representative agent knows the signal, i.e. the technology process, but the aggregation process makes the observations corrupted to the econometrician. Specifically we generate data from a small linearized DSGE model where decision rules are obtained under the assumption of  $AR(1)$  technology processes, but we simulate the time series with a fractional noise in the technology shock. This formulation implies a fractional  $AR(1)$  representation for the output which is in line with the results in Diebold and Rudebusch (1989) and Sowell (1992). Since the first order condition on labour choice is static and we want to isolate the effects of long memory on the estimation of the endogenous and exogenous states we employ a simple Ramsey model, featuring capital ( $k_t$ ), consumption ( $c_t$ ), output ( $y_t$ ), productivity ( $a_t$ ) and the real interest rate ( $r_t$ ).

The model can be described as a standard market problem as follows. Households choose consumption and investments such as to maximize an objective function:

$$U_0 = E_0 \sum_{t=0}^{\infty} \beta^t \{\log(c_t)\}, \quad (34)$$

subject to a budget constraint :

$$c_t + i_t \leq w_t + r_t k_{t-1}, \quad (35)$$

where  $w_t$  is the real wage rate; since household face no leisure choice, we normalize labour to one.

Capital is set by the households with the standard law of motion

$$k_t = (1 - \delta)k_{t-1} + i_t,$$

where no growth trend in productivity is assumed. Using eq.35, capital accumulation



can then be rewritten as:

$$k_t = (1 - \delta)k_{t-1} - c_t + w_t + r_t k_{t-1}, \quad (36)$$

The first necessary order condition associated with the maximization of the objective 34 subject to 36 is given by the standard Euler equation:

$$c_t^{-1} = E_t [\beta(1 + r_{t+1} - \delta)c_{t+1}^{-1}] \quad (37)$$

Firms use capital according to the production function:

$$y_t = a_t k_{t-1}^\alpha,$$

where technology evolves as:

$$\log(a_t) = \rho \log(a_{t-1}) + \epsilon_t,$$

where  $\epsilon_t$  is considered as i.i.d. by the households with a constant variance  $\text{var}(\epsilon_t) = \sigma_\epsilon^2$ .

Firms rent capital  $k_t$  by paying a rental price  $r_t$  and maximize their profits by choosing capital such that the real interest rate equals the marginal productivity of capital minus the depreciation rate  $\delta$ :

$$r_t + \delta = \alpha a_t k_{t-1}^{\alpha-1}$$

Finally we have the goods market clearing condition

$$y_t = c_t + i_t.$$

The competitive equilibrium for the economy is the sequence of prices  $\{r_t, w_t\}_0^\infty$  and quantities  $\{y_t, k_t, c_t\}_0^\infty$  such that firms maximize profits, agents maximize utility and all markets clear. Parameters are set in a way which is consistent with the literature. The capital share  $\alpha$  is set equal to 0.33; the preference term  $\beta$  is equal to

0.99 which corresponds to a real interest rate of 0.04 on annual basis; the depreciation rate  $\delta$  is set equal to 0.025; the autoregressive term is set to 0.9. To gain on clarity  $\sigma_\epsilon^2$  is set equal to 1.<sup>12</sup> The model is log-linearized around its steady state<sup>13</sup> and solved under the assumption that the innovations  $\epsilon_t$  are i.i.d. The reduced form solution of the model is written in the state space representation described in the previous section, namely

$$\begin{aligned}\theta_{t+1} &= \Phi\theta_t + \varepsilon_{t+1} \\ \mathbf{y}_t &= H\theta_t + \mathbf{v}_t\end{aligned}$$

where the vectors of state and observable variables as deviations from their steady state values are respectively given by

$$\begin{aligned}\theta_t &= [\ln(a_t/a^{ss}) \ln(k_t/k^{ss})] \\ \mathbf{y}_t &= [\ln(c_t/c^{ss}), \ln(y_t/y^{ss}), \ln(r_t/r^{ss})]\end{aligned}$$

where  $\varepsilon_{t+1} = [\varepsilon_{t+1}, 0]$ .

In order to remove any singularity in the system of equations we add in  $\mathbf{v}_t$  two i.i.d. measurement errors with zero mean and standard deviation equal to 0.01. Finally, we simulate the state space model to generate artificial data for consumption, output and the real interest rate, and generate long memory in the data by assuming that the productivity shock  $\varepsilon_t$  is a fractional noise,

$$(1 - L)^d \varepsilon_{t+1} = e_{t+1}$$

where  $e_t$  is a white noise. For any degree of fractional order  $d$  ranging from zero to one we generate 1000 samples of 170 observations respectively for output, consumption and the real interest rates. Then, using the Kalman filtering algorithm described in eq.21-29, for each sample we estimate through Maximum Likelihood the structural

---

<sup>12</sup>We checked that results are unchanged by setting lower values.

<sup>13</sup>The steady state values of the variables are denoted with <sup>ss</sup>

parameters and the implied dynamics for the state variables. Table one reports the simulation results and shows the true model parameters together with the mean of each estimated parameter for different degree of persistence. The last two rows report the Root Mean Square Error of the state variables ( $\theta_{t|t}$  and  $K_{t|t}$ ) predicted by the Kalman filter.<sup>14</sup>

Parameters	Fractional Integration Parameter $d$									
	True	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\alpha$	0.330	0.3478	0.3556	0.3645	0.3704	0.3702	0.3768	0.4109	0.4605	0.5483
$\delta$	0.025	0.0266	0.0272	0.0279	0.0279	0.028	0.0285	0.0309	0.0343	0.0407
$\beta$	0.99	0.9875	0.988	0.9881	0.991	0.9901	0.9903	0.9902	0.9923	0.9954
$\phi$	0.9	0.8903	0.8967	0.9004	0.9066	0.9067	0.9078	0.9126	0.9226	0.9388
$\sigma_e$	1	1.0404	1.0404	1.1358	1.2554	1.4744	1.703	1.8521	1.9403	1.9852
$\sigma_{v_1}$	0.01	0.1122	0.1384	0.0898	0.0698	0.0344	0.0258	0.0211	0.0211	0.0218
$\sigma_{v_2}$	0.01	0.1064	0.1314	0.1228	0.1289	0.1237	0.1359	0.2217	0.3609	0.6956
$RMSE \theta_{t t}$		1.0439	1.0967	1.1964	1.3862	1.7154	2.4849	5.6623	11.1167	19.6298
$RMSE K_{t t}$		0.3827	0.4714	0.5357	0.6491	0.8617	1.5441	5.3139	10.8163	19.0695

Table 1: Parameter estimation for different degrees of fractional integration  $d$

The bias tends to increase for all the model parameters as the persistence becomes stronger. For  $d$  bigger than 0.5, (Diebold and Rudebusch again) the bias becomes quite large especially for the capital share  $\alpha$ , the capital depreciation  $\delta$  and the variance of the productivity innovation  $\sigma_e$ : the estimated parameters are almost twice as large as the true parameters. This is fairly consistent with what we find by comparing the estimates of the augmented and the standard filter on US data in the next section. In fact, although the “true” parameters used in the simulation can be considered to be close to the ones adopted in the calibration literature, the estimation with U.S. data produces values which are significantly higher and consistent with a parameter  $d$  larger than 0.5.

<sup>14</sup>For each simulation sample we estimate the structural parameters of the models

Finally the RMSE of the state prediction for both capital and productivity (last two rows of Table one) are very large. This is consistent with the large estimates of variances of the measurement errors  $\sigma_{v_1}$  and  $\sigma_{v_2}$ . In fact, with a positive bias on the variances of the measurement errors, the estimated Kalman gain  $K$  is lower than the true Kalman gain, producing consequently a prediction of the state which is too smooth and insensitive to innovations compared to the real state. This reflects the inability of the model to capture the true persistence of the data which is discharged as a noise component.

In Table 2 we report the results of the same exercise for the augmented filter.<sup>15</sup>

Parameters	Fractional Integration Parameter $d$									
	True	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\alpha$	0.330	0.3302	0.3301	0.3301	0.3299	0.3299	0.3299	0.3309	0.3342	0.3411
$\delta$	0.025	0.025	0.025	0.025	0.0249	0.025	0.02499	0.0251	0.0253	0.0258
$\beta$	0.99	0.99	0.99	0.99	0.99	0.99	0.9900	0.99	0.9901	0.9903
$\phi$	0.9	0.90	0.90	0.90	0.90	0.9	0.899	0.9002	0.9007	0.902
$\sigma_e$	1	0.9964	1.0017	1.0106	1.01	1.0295	1.0437	1.0598	1.0836	1.1301
$\sigma_{v_1}$	0.01	0.0099	0.0099	0.0099	0.01	0.0099	0.0099	0.0099	0.0099	0.0099
$\sigma_{v_2}$	0.01	0.0099	0.0099	0.0099	0.01	0.0099	0.0100	0.0119	0.0249	0.0524
$RMSE \theta_{t t}$		1.0154	1.0172	1.0342	1.0654	1.1241	1.2155	1.3512	2.6863	5.3401
$RMSE K_{t t}$		0.2073	0.2542	0.3135	0.4014	0.5165	0.6899	0.9249	2.5774	5.1981

Table 2:

As it can be readily seen, the filter is able to estimate consistently the true parameters of the model. This is true for all the degree of fractional integration, even when the state variables become non-stationary ( $d > 0.5$ ). Not surprisingly, the augmented filter provides much more accurate prediction of the underlying state variables compared to the standard filter. In fact, the RMSE errors of the state

<sup>15</sup>We choose a value of  $m$ , number of lags in productivity, equal to 30. This seems to be a reasonable compromise; while implementing a rather effective correction it does not exclude too many observations.

prediction are up to 4 times smaller than those obtained with the standard filter. This result is quite important and shows the ability of the modified filter to capture the dynamic properties of the data. As we show in the last section, this accuracy in predicting the data dynamics also holds when using the modified filter to forecast out-of-sample the observed variables.

## 4 Real data estimation

### 4.1 Model

In this section we take our model to the real data and repeat the Maximum Likelihood estimation of an RBC model as in Ireland (2004); the same type of model is also used by Ruge-Murcia (2007), in order to compare different estimation techniques. Households choose consumption and labour/leisure and save by investing in stocks of capital. Furthermore, they maximize the following utility function:

$$U_0 = E_0 \sum_{t=0}^{\infty} \beta^t \{ \log(c_t) + \gamma(1 - n_t) \}, \quad (38)$$

subject to the budget constraint:

$$c_t + i_t \leq w_t n_t + r_t k_{t-1},$$

there is no population growth and the total amount of labour is normalized to one and leisure is given by  $1 - n_t$ . Capital accumulates with the standard law of motion

$$\eta k_t = (1 - \delta)k_{t-1} + i_t, \quad (39)$$

expressed in efficiency units in order to take into account a log-linear trend in technology  $\eta$ . This gives rise to the (standard) first order conditions:

$$c_t^{-1} = E_t [\beta(\eta + r_{t+1} - \delta)] c_{t+1}^{-1}, \quad (40)$$

$$w_t = \gamma c_t \quad (41)$$

The production function is given by the Cobb-Douglas:

$$y_t = a_t (\eta^t n_t)^{1-\alpha} k_{t-1}^\alpha, \quad (42)$$

where technology evolves as:

$$\log(a_t) = (1 - \rho) \log(a^{ss}) + \rho \log(a_{t-1}) + \epsilon_t. \quad (43)$$

In equilibrium the real interest rate equals the marginal productivity of capital minus the depreciation rate  $\delta$ :

$$r_t + \delta = \alpha a_t n_t^{1-\alpha} k_{t-1}^{\alpha-1},$$

and the real wage rate equals the marginal productivity of labour:

$$w_t = (1 - \alpha) a_t n_t^\alpha k_{t-1}^\alpha.$$

The model is closed by the market clearing condition:

$$y_t = c_t + i_t. \quad (44)$$

The competitive equilibrium for the economy is the sequence of prices  $\{r_t, w_t\}_0^\infty$  and quantities  $\{y_t, k_t, c_t, n_t, i_t\}_0^\infty$  such that firms maximize profits, agents maximize utility and all markets clear. The model is linearized by using the Taylor expansion of the system of equations (42), (40), (44), (43), (39), (41) around the steady state of the model. Then we solve the model following Klein (2000) and rewrite the reduced form solution into the (augmented) state space representation described in section 2. As before, we define with  $y_t$  the vector of endogenous observable variables of the model and with  $\theta_t$  the vector of unobservable exogenous processes. Following Ireland (2004), we introduce in the state space model, three independent autocorrelated

measurement errors,  $\eta_t = [\eta_t^y, \eta_t^c, \eta_t^h]'$  that are assumed to evolve as AR(1) processes

$$\begin{bmatrix} \eta_t^y \\ \eta_t^c \\ \eta_t^h \end{bmatrix} = \begin{bmatrix} \rho_{\eta^y} & 0 & 0 \\ 0 & \rho_{\eta^c} & 0 \\ 0 & 0 & \rho_{\eta^h} \end{bmatrix} \begin{bmatrix} \eta_{t-1}^y \\ \eta_{t-1}^c \\ \eta_{t-1}^h \end{bmatrix} + \begin{bmatrix} \zeta_t^y \\ \zeta_t^c \\ \zeta_t^h \end{bmatrix}$$

with

$$E \begin{bmatrix} \zeta_t^y \\ \zeta_t^c \\ \zeta_t^h \end{bmatrix} \begin{bmatrix} \zeta_t^y \\ \zeta_t^c \\ \zeta_t^h \end{bmatrix}' = \begin{bmatrix} \sigma_{\zeta^y}^2 & 0 & 0 \\ 0 & \sigma_{\zeta^c}^2 & 0 \\ 0 & 0 & \sigma_{\zeta^h}^2 \end{bmatrix}.$$

Therefore we have  $\mathbf{y}_t = [y_t, c_t, h_t]$  and  $\theta_t = [k_t, a_t, i_t, \eta_t^y, \eta_t^c, \eta_t^h]$ .

The following state space representation is then used to estimate the model:

$$\begin{aligned} \theta_{t+1} &= \Phi \theta_t + \varepsilon_{t+1} \\ \mathbf{y}_t &= H \theta_t + \mathbf{v}_t \end{aligned}$$

where

$$\begin{aligned} \Phi &= \begin{bmatrix} p_{kk} & q_{ka} \\ 0 & F \end{bmatrix}; \quad F = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & \rho_{\eta^y} & 0 & 0 \\ 0 & 0 & \rho_{\eta^c} & 0 \\ 0 & 0 & 0 & \rho_{\eta^h} \end{bmatrix} \\ H &= \begin{bmatrix} m_{ck} & n_{ca} \\ m_{yk} & n_{ya} \\ m_{rk} & n_{ra} \end{bmatrix} \end{aligned}$$

where the elements of the matrices  $\Phi$  and  $H$  are obtained from the solution of the model, i.e. for example:

$$k_t = p_{kk}k_{t-1} + q_{ka}a_t,$$

$$c_t = m_{ck}k_{t-1} + n_{ca}a_t.$$

In the presentation we focus only on one special treatment of measurement errors: three autocorrelated but mutually uncorrelated measurement errors. This is consistent with Ireland (2004) where he shows that such specification, compared with one having correlated measurement errors, have the best out-of sample forecasting properties in spite of less plausible values of the deep parameters. One of our claim is that this kind of trade off between forecasting and parameter estimation tend to vanish when long memory is taken into account.

Measurements are hours worked, consumption and GDP, which are taken respectively from BLS data (Current Employment Statistics) and US NIPA national accounting: data run from 1948:1 to 2002:2.<sup>16</sup> We estimate the following set of structural parameters:  $\alpha$ ,  $\eta$ ,  $\gamma$ ,  $\delta$ ,  $\rho$ ,  $\rho_{\eta^y}$ ,  $\rho_{\eta^c}$ ,  $\rho_{\eta^h}$ ,  $\sigma_\epsilon$ ,  $\sigma_{\zeta^y}$ ,  $\sigma_{\zeta^c}$ ,  $\sigma_{\zeta^h}$  plus the level of technology  $a^{ss}$  which enters the steady state expressions of the variables. The estimated parameters are the capital share, the log-linear trend of technology, the parameter which pins down the amount of hours worked in steady state, the depreciation rate of capital, the persistence parameter of the technology and the measurement errors together with their standard deviations, and the discount preference term  $\beta$ . Differently from Ireland, we estimate the depreciation term  $\delta$  in order to show that long-memory correction can help in the identification of parameters that are sometimes found to be difficult to identify in the literature, and therefore they are calibrated.

## 4.2 Parameter estimates

In this section we discuss the parameter estimates for two different models,  $AKF$  refers to the augmented state space model, while  $KF$  to the standard Kalman filter.

---

<sup>16</sup>we use the same dataset as in Ireland which is based on the 1996 chained data



Par.	$\alpha$	$\delta$	$\rho_a$	$\eta$	$\gamma$	$\rho_{\eta_t^y}$	$\rho_{\eta_t^c}$	$\rho_{\eta_t^h}$	$\sigma_\epsilon$	$\sigma_{\zeta^y}$	$\sigma_{\zeta^c}$	$\sigma_{\zeta^h}$
AKF	0.2917	0.0246	.969	.0055	.0041	.998	-.804	.9995	.0048	0.0039	4e-4	.0062
KF	0.5189	0.0031	.999	.0057	.0036	.999	.999	.9990	.0092	.0036	1e-4	.0060

Table 3: Parameter estimates

The estimation results differ substantially between the two filters<sup>17</sup>. Our finding suggests that there is a substantial amount of autocorrelation left over in the residuals  $\varepsilon_t$ , which the standard Kalman filter does not account for, and consequently leads to biased parameter estimates. From Table 3 we see that the share of capital estimated with the standard filter almost doubles the one given by the modified filter. There is a substantial difference in the values of both the persistence and variance parameters of the technology shock.

Overall, it seems that the augmented filter produces estimates which are in line with estimates produced by models with less restrictions on the variance-covariance matrices, such as the one with correlated shocks in Ireland (2004); moreover they are also closer to the values used in standard calibration exercises. Moreover there is a simple intuition explaining the large difference in estimated depreciation rates and capital shares. A large capital share and a low depreciation rate both make the policy function of capital more persistent. This happens both mechanically (less depreciation) and economically, since agents have more incentive, higher returns, to investing in capital goods. As a result the law of motion of capital in the reduced form of the model will have a stronger autoregressive term, i.e. it will be more persistent. Once data are cleaned from long memory, unexplained persistence does not bias estimates any longer.

<sup>17</sup>The same kind of result applies even when the  $\delta$  is calibrated

## 5 Forecasting

In order to assess the ability of our approach to capture the dynamic properties of the data we compare the out of sample forecast of the modified filter with those of the standard Kalman filter. The ratio behind this exercise follows the results in Granger and Joyeux (1980) who showed that although it is always possible to find an AR representation that can adequately fit long memory dynamics, the forecast produced by such AR models will not be very accurate.

The exercise is implemented as follows. We estimate the RBC model described in the previous section with both the standard and our approach for a subsample of data, precisely from 1948:1 until 1987:4. We generate out-of-sample forecasts one through four quarters ahead for each variable and compare the root-mean-squared forecast errors from the modified model to those from the standard Kalman filter. We then extend the subsample by one period and repeat the estimation and forecasting. We continue this way until all the sample is covered in 2002.

Table 4 reports the RMSE together for the forecasts generated by the standard Kalman filter (KF) and those generated by the augmented filter (AKF). In order to assess whether any difference of the two RMSE is significant we also report the statistic proposed by Diebold and Mariano (1995).<sup>18</sup>

---

<sup>18</sup>The critical values for the Diebold and Mariano test have been obtained using a bootstrap procedure.

Steps ahead	1	2	3	4
Output				
RMSE: AKF	0.6492	1.2294	1.8137	2.3585
RMSE: KF	0.984	1.8935	2.881	3.7871
D-M test	-6.8582***	-5.9845***	-4.8295***	-3.7659***
Consumption				
RMSE: AKF	0.4588	0.7022	1.0155	1.3689
RMSE: KF	0.5443	0.961	1.4172	1.9213
D-M test	-3.6916***	-3.7701***	-3.0075***	-2.1948**
Hours				
RMSE: AKF	0.6352	1.1373	1.6295	2.0847
RMSE: KF	0.8135	1.4583	2.0558	2.6461
D-M test	-4.8812***	-3.4673***	-2.2781**	-1.8388*

Table 4:

The results indicate that forecasts from the modified filter significantly outperform those from the standard Kalman filter. In particular, for output, the RMSE of the augmented filter are up to 60% smaller than those of the standard filter. This shows the better performance of the modified filter with respect to the normal one<sup>19</sup>. On one hand, the forecast from the augmented filter are more accurate than those from standard filter for all the steps ahead; on the other, the forecast accuracy of the modified filter improves, relatively to the standard filter, as the forecast horizon arises. This result indicates the presence of very persistent dynamics in the data that can not be fully captured by the autoregressive structure of the standard filter. On the other hand, it also shows that by accounting for such persistence it is possible to considerably improve the dynamic properties of the model.

<sup>19</sup>Similar, albeit less striking, results hold for the case when we calibrate the  $\delta$ .

## 6 Conclusion

In this paper we have shown a simple method to bring general equilibrium dynamic models to persistent data. This can be done by introducing one (or more) common factor long memory component as the driving forces of the data generating process. Existing techniques already allow researchers to estimate DSGE models under the assumption that data are either  $I(1)$  or  $I(0)$  with deterministic trend. In particular when data are  $I(1)$  it is possible to use a Diffuse Kalman filter under the assumption that the driving force of the economy such as technology is integrated of order one. This provides Maximum likelihood estimates of the deep parameters of the model under the assumption that there is a stochastic common trend.

A careful inspection of the data as well as relevant contributions to the econometric literature indicates that long memory is a relevant feature of real world economies. Long memory can be introduced in a general equilibrium framework by the aggregation of heterogeneous stochastic processes, as in Abadir and Talmain (2002), at the price of having a very large and hard-to-handle framework. In this paper we take the shortcut of following the RBC thread by assuming that technology is itself responsible for the long memory dynamic nature of the data. We believe to have good reasons for doing that since it is well documented by independent and methodologically different studies that productivity can have an  $I(d)$  nature.

We have shown that when unobservable states such as technology exhibit long range dependence, the standard Kalman filter approach to DSGE estimation is not well suited to tackle the estimation problem. In fact, the KF procedure can be summarized in two components: a state updating part and a signal extraction part (information set updating). Due to dynamic miss-specification, the direct updating of the states is incorrect; this generates a bias in the updating for the conditional variance of the unobservables. Due to the same miss-specification problem the orthogonality condition between projected states and the innovations constructed by the filter fails to hold. This affects the signal extraction part and breaks down the optimality of the recursive signal extraction method. We have found that the estimated deep parameters suffer of a bias; the estimation process and the filter incorrectly in-

troduces measurement error even when there is an almost zero measurement error in the data generating process.

We have proposed a method to take long memory into account. We view our procedure as a shortcut of the following assumption, that agents know the underlying dynamics of the economy and therefore their decision rules are correctly formed, still the econometrician cannot see individual shocks and its vision is corrupted by long memory and persistent autocorrelation. This informational assumption may be viewed as quite radical but we leave it to be relaxed in further research. Our work can also be viewed as a generalization of the methodology adopted in order to deal with trends: we filter the technology shocks in order to clean the state vector from long memory; this ensures that shocks are  $AR(1)$  with no more information left over in the residuals. The filtering technique we use is of a GLS type: we estimate the variance covariance matrix of the state vector, we factor its inverse by Cholesky decomposition and we use that to clean the state vector and to report it to an  $AR(1)$  with white noise.

Not only are the full sample estimates of our modified filter more in line with the calibration experience; we also show that in an out of sample forecasting exercise we outperform the conventional filter over all forecast horizons. This is due to the fact that we correctly take into account the autocorrelation of the data. Due to the flexible approach of Abadir and Talmain (2005) we evaluate the dynamic correlation of the state vector in such a way to be adequate even in the presence of a nonlinear autocorrelogram, while being able to approximate the (linear) fractional integration framework.

While we have applied our method to a simple RBC framework in principle there is no reason to be confined to such a unidimensional shock case. Several shocks can be accommodated in our framework. Nevertheless, as already said in the introduction, there is a widespread consensus about the  $I(d)$  nature of inflation; this would be a call for monetary and New Keynesian Models. Since these are mostly estimated by Bayesian Techniques we leave it to further research to extend our framework to that case.

## References

- [1] Abadir, K. M. and Talmain, G. (2002), “Aggregation, persistence and volatility in a macromodel”, *Review of Economic Studies*, 69, 749-779.
- [2] Abadir, K. M. and Talmain, G. (2005), “Distilling co-movements from persistence macro and financial series”, ECB working paper no. 525.
- [3] Abadir, K. M., Caggiano, G. and Talmain, G. (2006), “Nelson and Plosser revised: the ACF approach”, working paper 2005-7, University of Glasgow.
- [4] Altissimo, F., Mojon, B., Zaffaroni, P. (2005), “Fast micro and slow macro: can aggregation explain the persistence of inflation?”, ECB Inflation Persistence Network Paper.
- [5] Carvalho, C. (2006). ”Heterogeneity in Price Stickiness and the Real Effects of Monetary Shocks”, *The B.E. Journal of Macroeconomics: Vol. 2 : Iss. 1 (Frontiers)*
- [6] Chamber, M.J. (1998). ”Long memory and aggregation in Macroeconomic time series”, *International Economic Review*, 39 (4), 1053-1072
- [7] Diebold, F. X. and Rudebusch, G. D. (1989), ”Long memory and persistence in aggregate output”, *Journal of Monetary Economics*, vol. 24(2), 189-209.
- [8] Durbin, J. and Koopman, S.J. (2001), ”Time series analysis by state space methods”, Oxford University Press.
- [9] Gadea, M. and Mayoral, L. (2006), ”The Persistence of Inflation in OECD Countries: A Fractionally Integrated Approach,” *International Journal of Central Banking*, vol. 2(1).
- [10] Granger, C. W. J. (1980), “Long memory relationship and the aggregation of dynamic models”, *Journal of Econometrics*, 14, 227-238.

- [11] Granger, C. W. J. and Joyeux, R. (1980), An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-30.
- [12] Hamilton, J. (1994), *Time Series Analysis*. Princeton University Press.
- [13] Ireland, P. (2004), "A method for taking models to the data," *Journal of Economic Dynamics and Control*, vol. 28(6), pages 1205-1226.
- [14] Mayoral, L. (2005)., "Further evidence on the statistical properties of Real GNP," *Economics Working Papers* 955, Department of Economics and Business, Universitat Pompeu Fabra, revised Feb 2006.
- [15] McGrattan, E. (2007), "Measurement with Minimal Theory," 2006 Meeting Papers 338, Society for Economic Dynamics
- [16] McGrattan, E. (1994), "The macroeconomic effects of distortionary taxation," *Journal of Monetary Economics*, Elsevier, vol. 33(3), 573-601.
- [17] Moretti, G. (2007), "Detecting long memory co-movements in macroeconomic time series", Bank of Italy working paper 642.
- [18] Ravenna, F. (2007), "Vector Autoregressions and Reduced Form Representations of DSGE models" *Journal of monetary economics*, Vol. 54 N. 7
- [19] Robinson, P. (2003), *Time Series with long memory*. Oxford University Press.
- [20] Ruge-Murcia, F. J. (2007). "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, vol. 31(8), 2599-2636.
- [21] Sowell, F. (1992), "Modeling long-run behavior with the fractional ARIMA model", *Journal of Monetary Economics* 29, 277-302.

Département des Sciences Économiques  
de l'Université catholique de Louvain  
Institut de Recherches Économiques et Sociales

Place Montesquieu, 3  
1348 Louvain-la-Neuve, Belgique